

Quick Job Entry or Long-Term Human Capital Development? The Dynamic Effects of Alternative Training Schemes

ADERONKE OSIKOMINU

University of Zurich, Albert-Ludwigs-University Freiburg, CESifo and IZA

First version received October 2009; final version accepted March 2012 (Eds.)

This article investigates how precisely short-term, job search-oriented training programs as opposed to long-term, human capital intensive training programs work. We evaluate and compare their effects on time until job entry, stability of employment, and earnings. Further, we examine the heterogeneity of treatment effects according to the timing of training during unemployment as well as across different subgroups of participants. We find that participating in short-term training reduces the remaining time in unemployment and moderately increases job stability. Long-term training programs initially prolong the remaining time in unemployment, but once the scheduled program end is reached participants exit to employment at a much faster rate than without training. In addition, they benefit from substantially more stable employment spells and higher earnings. Overall, long-term training programs are well effective in supporting the occupational advancement of very heterogeneous groups of participants, including those with generally weak labor market prospects. However, from a fiscal perspective only the low-cost short-term training schemes are cost efficient in the short run.

Keywords: Training, Program Evaluation, Duration Analysis, Dynamic Treatment Effects, Multiple Treatments, Active Labor Market Policy

JEL Codes: J64, C41, J68, I28

1. INTRODUCTION

Job training programs are the core of active labor market policy in most advanced countries. They aim at integrating the jobless and economically disadvantaged into the labor market. Countries like the USA and Germany have been operating training policies for many decades. Yet the success of different approaches—in particular, human capital development as opposed to work-first strategies—remains a controversial issue. Long-term, human capital intensive training schemes provide comprehensive instruction in occupational skills. While they focus on improvement in the productivity of the unemployed, they usually do not aim at rapid reemployment. Moreover, they are relatively expensive. In contrast, job search assistance programs comprise job readiness training and instruction in job search skills. They are based on the idea that occupational skills are best acquired on the job. Consequently, they focus on quick job entry. A major advantage is their low cost. This approach has been emphasized in policy reforms implemented in 1996 in the

USA and the late 1990s to early 2000s in other advanced countries.¹ However, the limited set of skills provided may not be sufficient to address structural mismatch and to improve employment stability in the long run, a compelling concern in view of sweeping shifts of labor demand that disadvantage less-skilled workers.² After the sharp downturn of the 2008/2009 recession, many economists call for an expansion of public sponsored training programs to combat increasing rates of long-term joblessness among less-skilled persons.³

What is the best design of job training for the unemployed and the economically disadvantaged? This article sets out to investigate how precisely short-term, job search-oriented training and long-term, human capital intensive training programs work. We examine and compare their effects on time until job entry, stability of employment, and earnings within a unified framework. For this purpose, we model the full path of transitions between different labor market states and training programs over time. Our dynamic approach has several important advantages over conventional static research designs. First, we are able to net out differential changes in the composition of treatment and comparison group persons over time. This is necessary in order to separately assess the impacts of training on unemployment and employment spells and to obtain impact estimates on employment stability and wages that are not biased by systematic differences between treated and comparison persons who take up a job. Second, we are able to exactly align treated and comparison persons with respect to their prior unemployment experience and calendar time.⁴ Thus, we avoid that for example, differences in labor market conditions upon reemployment confound employment impacts of training. Third, our continuous-time framework avoids specification issues that arise as a consequence of aggregating along the time dimension.

A dynamic approach is also necessary because in many countries with comprehensive active labor market policies, like in our case of Germany, program assignment is not a static, one-time event. Rather, it is a recurring decision that is dynamically related to the success of job search. Job-seekers with longer unemployment spells are more likely to end up receiving treatment because they have been exposed to the assignment process for a longer period of time.⁵ We therefore explicitly specify the accumulation of information over time and model the dynamic process of treatment assignment jointly with the transitions between labor market states. This strategy allows us, in addition, to analyze the effect of the timing of program participation during unemployment—an important dimension of effect heterogeneity that has been neglected in the literature so far.

Further, we investigate the heterogeneity of treatment effects not only according to the timing of training during unemployment but also across different subgroups of participants. In particular, we examine how different occupational groups fare with the two training programs. This allows us to address the question whether job training programs are an effective tool to mitigate the

1. See e.g. Blank (2002); European Commission (2002); OECD (2005, ch. 4; 2006) and section B.2 in the Supplementary Appendix.

2. See e.g. Acemoglu and Autor (2011), Autor, Katz, and Kearney (2008), Dustmann, Ludsteck, and Schönberg (2009) and Spitz-Oener (2006) for evidence on long-term shifts in the structure of wages and employment especially in the USA and Germany.

3. See e.g. Elsby, Hobijn, and Şahin (2010) and the accompanying comments as well as the article “America’s Jobless Men—Decline of the Working Man” in *The Economist*, 28 April, 2011.

4. Ashenfelter (1978); Heckman, Ichimura, and Todd (1997); and Heckman and Smith (1999), among others, stress the importance to carefully align treated and comparison persons with respect to their pretreatment outcomes.

5. See Fredriksson and Johansson (2008), for a formal analysis of dynamic program starts. A similar situation arises if there is a delay between the moment of (random) assignment and program start. Some persons may drop out of the treatment group before the start of the intervention because they already found a job on their own, see Heckman, LaLonde, and Smith (1999) for evidence of substantial dropout rates in major US social experiments. Abbring and van den Berg (2005) formally analyze the value of randomization at some baseline point if actual enrolment in treatment is dispersed over time.

negative effects of structural labor demand shifts. We give a detailed picture of the dynamics of training impacts, provide summary statistics of overall impacts, e.g. in terms of mean and median outcomes, and conduct a cost-benefit assessment. We thus provide an exceptionally detailed account of these two widely used active labor market programs. Such evidence is important in order to optimally allocate existing training programs and to devise new, improved schemes.

To provide the conceptual backdrop for our empirical analysis, we introduce a novel dynamic potential outcome framework based on the theory of continuous-time counting processes. We formally describe the dynamic relationship between treatment times, potential, and actual outcome times based on the notions of past, present, and future. We formulate the assumptions needed to solve the evaluation problem in a dynamic setup in which the conditioning events are sequentially updated over time. We then develop our dynamic causal model as a semi-Markov process for the transitions between the different labor market states and training programs. In this way, we are able to extend the timing-of-events approach by Abbring and van den Berg (2003a) to a setting that involves sequential updating of information, multiple treatments (short-term and long-term training), multiple outcomes (unemployment duration, employment duration, and earnings), and treatment effects that depend in a general way on the past unemployment experience, the treatment time, observed covariates, and unobserved heterogeneity.

Our identification strategy relies on the conditional no-anticipation and independence assumptions as well as on results from the literature on mixed hazard rate models. Under the conditional no-anticipation assumption, the current unemployment outlook is the same for any two potential outcomes associated with different future treatment times conditional on time-constant unobservables, the elapsed unemployment experience, and the covariate history up to the current time. Under the conditional independence assumption, the current unemployment outlook of potential outcomes associated with different treatment times is conditionally independent of treatment decisions taken up to the current time. Taken together the two assumptions can be interpreted as conditional sequential randomization into treatment. The combination of three factors makes them a credible basis for causal analysis in our case. First, we focus on a homogenous sample of job-seekers with a continuous prior employment record. Second, we exploit that the allocation of training programs is driven by the short-term supply of training slots as well as private information of the caseworker. Third, we adopt a flexible modeling strategy that relies on large sample sizes, repeated observations per person and a rich set of time-constant as well as time-varying covariates.

Under conditional no-anticipation and independence, we can then study identification of the joint distributions needed for causal inference on training impacts. Specifically, we view the successive transitions from unemployment and employment as a sequence of competing and single risks models with lagged occurrence dependence and lagged duration dependence in the higher order transition rates. We apply results from the literature on duration models with proportionate unobserved heterogeneity in the hazard rates, dependent competing risks, and lagged duration dependence (e.g., Abbring and van den Berg, 2003b; Brinch, 2007; Honoré, 1993; Horny and Picchio, 2010). Hence, our estimated treatment effects are identified semiparametrically. Our large and exceptionally informative data allow us to dispense with some important functional form assumptions that are typically made in the literature and to adopt a rich and flexible modeling strategy. In particular, we exploit that we observe repeated realizations per person and time-varying covariates.⁶ Thus, we are able to overcome important limitations of existing evaluation studies using duration methods for program evaluation.

6. With these data features, identification only relies on separability of the hazard rate in the unobserved heterogeneity term.

The empirical analysis in this article uses unique administrative data for Germany. Germany is an interesting case to study because its recent reforms and developments in the field of labor market policy closely follow the recommendations formulated in the international policy debate in the mid-1990s in view of high unemployment levels especially in the European countries (cf. European Commission 2002, on the “European Employment Strategy” and OECD, 2006, on the “OECD Jobs Strategy”). Average public spending on active labor market policies per unemployed relative to per capita GDP in the years 2000 and 2001 is around 33% in Germany which is at the mean across OECD countries (OECD, 2006, ch. 7).⁷ Traditionally, long-term, human capital development-oriented training programs dominated active labor market policy in Germany. Since the turn of the millennium, short-term, job search-oriented training programs have gained in importance.⁸

Our main findings are as follows. Participating in short-term training reduces the remaining time in unemployment and has moderate positive effects on subsequent job stability. Long-term training programs initially prolong the remaining time in unemployment. However, once the scheduled program end is reached participants exit to employment at a much faster rate than without training. Moreover, they benefit from substantially more stable employment spells and higher earnings. Further, our findings point to the possibility of improving the efficiency of long-term training programs through a careful targeting. Specifically, we find that the opportunity cost of participating in terms of prolonged unemployment is lower for people with lower chances to exit unemployment on their own, e.g. long-term unemployed and low skilled. Persons without formal education degree and persons previously working in low- and medium-skilled manual occupations reap particularly high gains with respect to earnings. Persons previously working in medium-skilled analytic and interactive jobs show substantial gains in terms of employment stability. Overall, long-term training programs are well effective in supporting the occupational advancement of very heterogeneous groups of participants, including those with generally weak labor market prospects. However, from a fiscal perspective only the low-cost short-term training schemes are cost efficient in the short run. In light of the recent policy shifts favoring short, job search-oriented training programs a more balanced use of both types of training seems warranted. In fact, while the shorter programs may appeal for their cost efficiency the longer programs seem more effective in tackling structural deficits.

The remainder of the article is organized as follows. The next section discusses related literature in the field. Section 3 lays out our evaluation approach and Section 4 describes the implementation. Section 5 presents our causal estimates of training impacts, and Section 6 concludes. Sections A through I in the Supplementary Appendix contain further information on the identification analysis, the institutional context, the data source, descriptive evidence, the variables used in the estimation, and the estimated models.

2. RELATED LITERATURE

The literature on microeconomic evaluations of job training programs has been growing rapidly since the late 1990s as the political demand for rigorous scientific evaluations of labor market programs increased also in the continental European countries. Evidence reviewed in various surveys and meta-analyses e.g., by Heckman *et al.* (1999); Greenberg, Michalopoulos, and Robins (2003); Kluve (2010); and Card, Kluve, and Weber (2010) suggests that human capital intensive training programs show modestly positive effects in the medium run, with the percentage gain in

7. The corresponding figure is 7% for the USA, and 51% for Sweden.

8. See section B.2 in the Supplementary Appendix.

employment generally exceeding that in earnings. Job search assistance programs yield favorable employment effects in the short run. In the following, we highlight evaluation studies that are most closely related to ours.

We start with studies applying experimental or matching methods. The vast majority of evaluation studies analyzes short- to medium-term outcomes of job training programs, which may bias findings toward a more favorable assessment of shorter programs, such as job search training, compared with longer ones. A few studies investigate long-term impacts, among them Couch (1992) and Hotz, Imbens, and Klerman (2006) for the USA as well as Fitzenberger, Osikominu, and Völter (2008) and Lechner, Miquel, and Wunsch (2011) for Germany. Their results suggest that, although positive effects of human capital intensive training appear only with a delay, they persist over periods as long as eight or nine years. Hotz *et al.* (2006) and Dyke *et al.* (2006) study the differential effects of training programs aiming at human capital development and job search assistance, respectively, in the USA. They find that job search assistance programs improve employment prospects in the short run, whereas more intensive training programs initially lead to employment and earnings losses. Over the longer term, human capital intensive training tends to be more effective than job search assistance.

While experimental and matching estimators that closely mimic the experimental design are appealing for their straightforward interpretation, they also have some limitations. As argued above, the experimental setup is inherently static and does not take into account the dynamics of program assignment and of labor market outcomes.⁹ In order to better study the labor market dynamics associated with program participation some researchers have used event history models to evaluate training programs. The earlier duration literature—notably Gritz (1993), Ham and LaLonde (1996) and Eberwein, Ham, and LaLonde (1997) for the USA and Bonnal, Fougère, and Sérandon (1997) for France—focuses on modeling the dynamic selection into different labor market states. While in Ham and LaLonde (1996) training is randomly assigned to a stock sample of eligible persons, Gritz (1993), Bonnal *et al.* (1997), and Eberwein *et al.* (1997) jointly model transitions into program states and outcome states. This early duration literature conceives program participation as a separate labor market state and models treatment effects with an indicator for past program participation in subsequent transition rates. These studies find mixed effects on subsequent unemployment spells and mostly positive impacts on subsequent employment spells.

More recently, Abbring and van den Berg (2003a) develop an econometric framework to jointly analyze dynamically assigned programs and the probability of survival in a baseline state. They prove identification of dynamic treatment effects that depend on observed covariates and either the elapsed time since treatment or unobserved heterogeneity.¹⁰ Compared with the early duration literature the more recent timing-of-events approach explicitly models the dynamics of program starts. Assignment to program and exit from the baseline state are viewed as two competing risks. If exit from the baseline state occurs first, the waiting time until treatment is censored. Conversely, if assignment to treatment occurs first, the exit rate from the baseline

9. There exist a number of intermediate strategies between a fully dynamic approach and a completely static design. Matching estimators have been adapted to account for dynamic program assignment, see e.g. Sianesi (2004, 2008); Dyke *et al.* (2006); Fitzenberger *et al.* (2008). Other approaches combine matching with dynamic outcomes such as transition rates and survival times, see e.g. Bergemann, Fitzenberger, and Speckesser (2009) and de Luna and Johansson (2010). See Abbring and Heckman (2007), Abbring and van den Berg (2004), and Crépon, Ferracci, Jolivet and van den Berg (2009) for a methodological overview and comparison of different evaluation approaches.

10. Richardson and van den Berg (2006) extend the timing-of-events approach proving identification of dynamic treatment effects that are a proportional function of time since treatment, observed covariates and unobserved heterogeneity. Abbring (2008) develops a framework based on mixed proportional hazard rate models that involves multiple baseline states and competing destination states.

state may change as a consequence of the treatment. In addition, the timing-of-events approach highlights the importance of no-anticipation of future treatments for identification. Our dynamic causal model is similar to that of Abbring and van den Berg (2003a), but we use a counting process framework to describe the current evolution of potential outcome processes with respect to past and future treatment events. We adapt the conditional no-anticipation and independence assumptions to a setting in which the conditioning events are sequentially updated over time. Further, we consider multiple types of treatments and multiple outcomes as well as treatment effects that depend in a general way on the past unemployment experience, the treatment time, observed covariates, and unobserved heterogeneity.

A small literature applies the timing-of-events approach to evaluate occupational training (Crépon, Ferracci, and Fougère, 2012, for France; Lalive, van Ours, and Zweimüller, 2008, for Switzerland; Hujer, Thomsen, and Zeiss, 2006b, for Germany, and Richardson and van den Berg, 2006, for Sweden) or active job search programs (Crépon, Dejemeppe, and Gurgand, 2005, for France; Hujer, Thomsen, and Zeiss, 2006a, for Germany; and Weber and Hofer 2004a,b, for Austria). The vast majority focuses on impacts on the duration in the initial unemployment spell, only Crépon *et al.* (2005) and Crépon *et al.* (2012) also study unemployment recurrence. The evidence in these studies suggests that job search training reduces unemployment duration, while more intensive training programs tend to increase it. Crépon *et al.* (2005) and Crépon *et al.* (2011) find that both program types have a beneficial effect on unemployment recurrence.

A major advantage of our study is its exceptionally large and informative administrative data. Thus, we can relax important functional form assumptions that are typically made in the duration literature. We adopt a flexible modeling strategy and provide a detailed account of job search-oriented and human capital intensive training with respect to the outcomes unemployment duration, employment duration, and earnings. As a further contribution, we examine the heterogeneity of treatment effects across a range of important observed characteristics such as education and occupation.

3. EVALUATION APPROACH

3.1. *Dynamic potential outcome framework*

Consider the following setup. People can either be unemployed or employed. While unemployed, they may be assigned to a training program. In particular, the job-seeker and the caseworker at the local employment agency meet repeatedly during the unemployment spell. At any such occasion, the caseworker decides whether to assign the job-seeker to a program or to postpone participation to the future, waiting further how job search evolves. Somebody who has not participated, say, until Day 80 of his unemployment spell may still enrol later. If, however, he starts a new job at Day 81 he would not be eligible for participation anymore. Thus, a job-seeker in open, i.e., untreated, unemployment is exposed to two risks that compete to end open unemployment: start of a training program and start of a new job. How can we evaluate the effect of participating in training on the duration of unemployment in this setting?

We adopt a continuous-time version of the potential outcome approach to program evaluation (Neyman, 1923; Roy, 1951; Rubin, 1974).¹¹ The different starting times of training during unemployment represent a continuum of mutually exclusive treatments $s \in \mathbb{R}_+ \cup \{\infty\}$. $s = \infty$ corresponds to the case that training never starts. We denote by $T^*(s) \in \mathbb{R}_+$ the potential duration in unemployment until exit to employment that prevails if training starts at time s of

11. See Abbring and van den Berg (2003a) and Lok (2008) for related approaches in continuous time. Abbring and Heckman (2007) provide an overview of different dynamic evaluation approaches.

unemployment.¹² Let $T \in \mathbb{R}_+$ be the actual unemployment duration. For training starting at s , the actual outcome time is identical to the potential outcome time associated with s , i.e., $T \equiv T^*(s)$, whereas the other potential outcome times $T^*(s')$, $s' \neq s$, are counterfactual.

We express the random time spent in unemployment as a continuous-time stochastic process. This will allow us to formally describe how past and future treatment times affect the current unemployment outlook conditional on the information accumulated until the current period. Let u denote the elapsed time since the beginning of unemployment. The actual unemployment duration T generates a counting process $Y \equiv \{Y(u), u \in \mathbb{R}_+\}$ such that $Y(u) = 0$ if $u \leq T$ and $Y(u) = 1$ if $u > T$. Analogously, we denote the counting process associated with the potential unemployment duration $T^*(s)$ as $Y^*(s) \equiv \{Y^*(s, u), u \in \mathbb{R}_+\}$ such that $Y^*(s, u) = 0$ if $u \leq T^*(s)$ and $Y^*(s, u) = 1$ if $u > T^*(s)$.¹³ Further, let $dY(u)$ be the differential of $Y(u)$ in a small interval $[u, u + du)$. Let u^- denote the time just before u . The probability that an exit to employment occurs in the small interval $[u, u + du)$ conditional on survival in unemployment until just before u is given by $\Pr(dY(u) = 1 | Y(u^-) = 0) = \Pr(T \in [u, u + du) | T \geq u) \equiv \theta(u)du$, where $\theta(u)$ denotes the hazard rate or intensity process. In order to model this probability conditional on the information available up to the current period, we introduce the filtration $\mathcal{I}(u^-)$ that collects the information accumulated until just before u . Let $\{x(u), u \in \mathbb{R}_+\}$ be a vector-valued stochastic process of time-varying random variables and v a vector of time-constant random variables.¹⁴ While $x(u)$ represents covariates that can be observed by both the job-seeker and the researcher, v is not observable for the researcher. Now $\mathcal{I}(u^-)$ is defined as $\sigma\{Y(r), x(r), v, 0 \leq r < u\}$, where $\sigma\{\cdot\}$ is the σ -algebra generated by the events $\{Y(r), x(r), v, 0 \leq r < u\}$. Y has an intensity process with respect to the filtration $\mathcal{I}(u^-)$ if $\theta(u | \mathcal{I}(u^-))$ is predictable given $\mathcal{I}(u^-)$ and, thus, the conditional expectation of $dY(u)$ equals the conditional intensity process times the increment of u : $E[dY(u) - \theta(u | \mathcal{I}(u^-))du | \mathcal{I}(u^-)] = \Pr[dY(u) = 1 | \mathcal{I}(u^-)] - \theta(u | \mathcal{I}(u^-))du = 0$.

Our goal is to contrast the marginal distributions of potential outcome times associated with different treatment times. In particular, we want to evaluate the effect of starting training at s instead of s' for those who enrol at s . The potential outcome distribution associated with starting training at s' is counterfactual for those who enrol at s . In order to solve the evaluation problem we rely on two assumptions. As only unemployed persons are eligible for training the treatment times are censored for people who find a job before starting training. Therefore, we assume that conditional on the information accumulated up to just before u the probability of an exit to employment in a small interval $[u, u + du)$ is the same across potential outcomes associated with different treatment times that lie after u . Formally, for any two potential outcome processes $Y^*(s)$ and $Y^*(s')$ that have intensity processes with respect to the filtration $\mathcal{I}(u^-) \equiv \sigma\{Y(r), x(r), v, 0 \leq r < u\}$, we suppose that:

$$(Conditional\ No-Anticipation) \tag{1}$$

$$\Pr[dY^*(s, u) = 1 | \mathcal{I}(u^-)] = \Pr[dY^*(s', u) = 1 | \mathcal{I}(u^-)] \quad \forall u \leq \min(s, s').$$

Under the conditional no-anticipation assumption (1), the current probabilities of survival in unemployment for any two potential outcome processes associated with different future treatment times coincide. Hence, we can study the distribution of the observed unemployment process given

12. We assume that all potential outcomes $\{T^*(s), 0 \leq s \leq \infty\}$ are absolutely continuous random variables.

13. We assume that all counting processes are right-continuous and have limits on the left.

14. We suppose that the process $\{x(u), 0 \leq u\}$ is left continuous and that v is bounded.

that training has not yet started to recover the potential unemployment process without training ($s = \infty$).¹⁵

Second, we assume that conditional on the history of information available just before u the potential outcome processes are independent of the treatment history up to u . Formally, the random waiting time in unemployment until the start of training induces a counting process $\mathbf{D} \equiv \{D(u), u \in \mathbb{R}_+ \cup \infty\}$ such that $D(u) = 0$ if u is smaller or equal to the treatment time and $D(u) = 1$ else. Denote the σ -algebra generated by the treatment process as $\sigma\{D(r), 0 \leq r < u\}$. We suppose that the potential outcome process $Y^*(s)$ has an $\mathcal{I}(u^-)$ -predictable intensity process that is also the intensity process of $Y^*(s)$ with respect to the extended filtration $\mathcal{J}(u^-) \equiv \mathcal{I}(u^-) \cup \sigma\{D(r), 0 \leq r < u\}$, i.e.:

$$(Conditional\ Independence) \quad (2) \\ \Pr[dY^*(s, u) = 1 | \mathcal{J}(u^-)] = \Pr[dY^*(s, u) = 1 | \mathcal{I}(u^-)] \quad \forall s.$$

While assumption (1) characterizes the evolution of current potential outcomes associated with future treatment times, assumption (2) refers to the relationship between current potential outcomes and past treatment events. Taken together the two assumptions can be interpreted as sequential randomization into treatment conditional on the information available up to the time of randomization. They are comparable with sequential conditional unconfoundedness assumptions invoked in sequential matching approaches (e.g., Lok, 2008; Robins, 1998; Sianesi, 2004), except that we allow the conditioning set to include time-constant unobservables. Importantly, assumptions (1) and (2) do not rule out that job-seekers can in general predict the probability of receiving training at particular points in time. Moreover, their predictions may be based on information acquired during the course of unemployment. Assumptions (1) and (2) just require that the researcher can control for all events that jointly predict treatment and outcome times. In addition, after conditioning on all relevant information, treatment assignment still needs to be a stochastic process.¹⁶ In Section 3.3, we substantiate the empirical support of assumptions (1) and (2) in our application.

By assumptions (1) and (2), we can link the potential probability of an exit to employment at time u to the actually observed probability:

$$\Pr[dY^*(s, u) = 1 | \mathcal{I}(u^-)] = \Pr[dY^*(s, u) = 1 | \mathcal{J}(u^-)] = \Pr[dY(u) = 1 | \mathcal{J}(u^-)].$$

Thus, we can conduct inference on the potential unemployment processes associated with different treatment times by studying the properties of the actually observed unemployment processes conditional on the respective treatment times. Further, we can express the actual outcome process and the treatment process as three underlying processes that are independent

15. Without the conditional no-anticipation assumption, we could only identify the average of pretreatment potential outcomes with respect to the distribution of future starting dates.

16. Abbring and van den Berg (2003a) also rely on versions of the conditional no-anticipation and independence assumptions. Translated into our notation, their no-anticipation assumption states that $\Pr[dY^*(s, u) = 1 | \mathcal{H}(u^-)] = \Pr[dY^*(s', u) = 1 | \mathcal{H}(u^-)] \quad \forall u \leq \min(s, s')$, where $\mathcal{H}(u^-) \equiv \sigma\{Y(r), 0 \leq r < u, x(0), v\}$. Thus, they abstract from the possibility that the current transition probabilities of potential outcomes associated with different future treatment times depend differentially on events occurring between the start of unemployment and the current period. Translated into our notation, their conditional independence assumption states that the potential outcome processes, $\{Y^*(s), s \in \mathbb{R}_+ \cup \infty\}$, are independent of the treatment process, \mathbf{D} , conditional on the information available at time zero, $\mathcal{I}(0) \equiv \sigma\{x(0), v\}$. This assumption abstracts from the possibility that events occurring during the course of unemployment may jointly influence potential outcome times and treatment times.

conditional on $\mathcal{J}(u^-)$. The first component process tracks the transition from open unemployment to employment, the second the transition from open unemployment to treatment and the third the transition from treated unemployment to employment. Thus, the waiting times in open unemployment until the start of training and until exit to employment are two competing risks. Identification of the joint distribution of the waiting times is nontrivial because we allow all three to depend on the unobserved heterogeneity v . We rely on results from the literature on mixed hazard rate and competing risks models. We describe the full model, that includes two types of training programs, and its identification in Section 3.2.

Further, we want to study the impact of training on the stability of subsequent employment. We extend our framework as follows. Let $K(u)$ be the state occupied at time u since inflow into unemployment. $K(u) = U$ if unemployed and $K(u) = E$ if employed at time u . The transition from origin state k to destination state l , can be modeled as a continuous-time semi-Markov process. Analogously to above, let $T_{kl} \in \mathbb{R}_+$ denote the random time spent in state k until exit to state l and Y_{kl} the corresponding counting process that tracks the transition from state k to state l . Similarly, $Y_{kl}^*(s)$ denotes the potential outcome process for the transition from state k to l prevailing if training starts at time s of unemployment. The probability that a transition from k to l occurs in a small interval $[u, u + du)$ is given by $\Pr[dY_{kl}(u) = 1 | Y_{kl}(u^-) = 0, K(u^-) = k] = \Pr[K(u + du) = l | T_{kl} \geq t, K(u^-) = k]$, where t is the elapsed duration in the current state k .

With some small modifications, assumptions (1) and (2) carry over to the extended setup. In both equations, substitute Y_{kl}^* for Y^* . The filtrations are now defined as $\mathcal{I}(u^-) \equiv \sigma\{Y_{kl}(r), x(r), v; 0 \leq r < u; k, l \in \{E, U\}, k \neq l\}$ and $\mathcal{J}(u^-) \equiv \sigma\{Y_{kl}(r), D(\min(r, T_{UE})), x(r), v; 0 \leq r < u; k, l \in \{E, U\}, k \neq l\}$. Further, we assume that the conditional no-anticipation assumption (1) holds for all $u \leq \min(s, s')$ if $K(u) = U$ and for all $s, s' > T_{UE}$ if $K(u) = E$. Thus, the survival experience is the same for potential employment processes associated with different treatment times that have not been realized during the preceding unemployment spell. Under assumptions (1) and (2), we can recover the potential employment processes associated with different treatment times from the actual employment process conditional on the respective treatment times. Hence, we view our dynamic causal model as a semi-Markov process that involves a competing risks model for the waiting times in open unemployment until exit to employment and training, and single risk models for the transitions from treated unemployment to employment as well as from treated and nontreated employment to unemployment. In addition, we consider a second model with earnings instead of employment duration. In this case, we model the conditional distribution of earnings as a hazard rate together with the hazard rates from unemployment into training and into employment.¹⁷ We describe our setup and the identification in detail in the next section.

3.2. Identification with proportionate unobserved heterogeneity

We now describe our empirical setup that comprises two treatments i.e., short-term training and long-term training, and two outcomes i.e., the duration until reemployment and the duration of subsequent employment or the level of subsequent earnings in terms of a continuous-time semi-Markov process. We then discuss identification of our model based on results from the literature on mixed hazard rate models with proportionate unobserved heterogeneity.

Consider an inflow sample into the initial state open unemployment O . A person in the initial state may exit to any of the three states (1) unemployment treated with short-term training S , (2) unemployment treated with long-term training L , or (3) employment E . A person in state

17. Donald, Green, and Paarsch (2000) also use a hazard rate model to investigate differences between the US and Canadian conditional wage distributions.

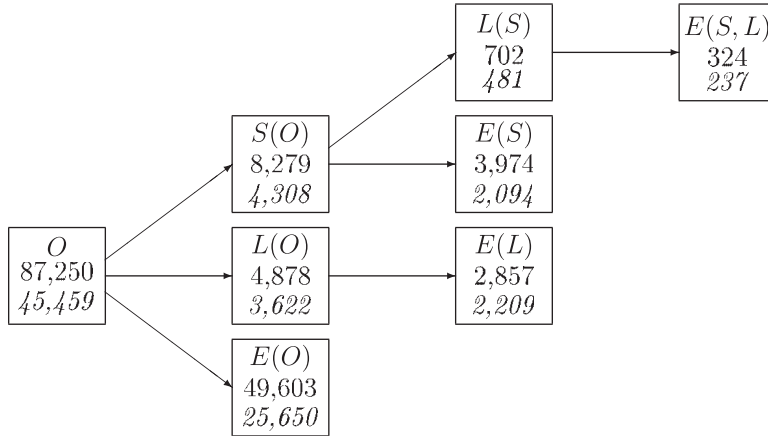


FIGURE 1

Number of observations at different phases of the unemployment-employment cycle.

Notes: O, S, L, E denote the labor market states open unemployment, unemployment treated with short-term training, unemployment treated with long-term training, and employment, respectively. The states in parentheses indicate dependence of current states on the past trajectory. The first number in each box refers to the number of observations using repeated cycles per person, the second number (in italics) to the number of observations using only the first cycle.

S may exit to the destinations L or E . A person in state L may move to state E .¹⁸ We call one such sequence of transitions from unemployment to employment, a cycle. In the model with employment duration as the second outcome, we also model the transition from employment, E , back to open unemployment, O , and consider repeated cycles per individual. In the alternative model with earnings as the second outcome, we use just the first cycle for every individual.¹⁹ Figure 1 illustrates the possible pathways from open unemployment to employment.

More formally, let $\mathcal{K} = \{O, S, L, E\}$ denote the state space. The function $\mathcal{Z}(k)$ assigns to each origin state $k \in \mathcal{K}$ a set of possible destination states:

$$\mathcal{Z}(k) = \begin{cases} \{S, L, E\} & \text{if } k = O \\ \{L, E\} & \text{if } k = S \\ \{E\} & \text{if } k = L \\ \{O\} & \text{if } k = E. \end{cases}$$

For $k \in \{O, S\}$, the number of elements in $\mathcal{Z}(k)$ is greater than one, $|\mathcal{Z}(k)| > 1$. Thus, we view a cycle as a sequence of competing risks and single risk models.

According to the Markov assumption, the probability to experience a transition from current state k to destination state l only depends on the current state k occupied. In order to differentiate e.g., between employment spells following a participation in training as opposed to open unemployment (cf. figure 1), we define the augmented state space $\mathcal{K}^a = \{O, S(O), L(O), E(O), L(S), E(S), E(S, L)\}$, where the states in parentheses indicate the past

18. If short-term training is the first intervention we allow for a second treatment with long-term training because short-term training is also used to assess the professional skills of an unemployed and to define a suitable reintegration plan, which may entail a later participation in a long-term training program.

19. Due to the reporting rules of the employment register we cannot model transitions into and out of employment at the same time as earnings, see Section 4.2.

trajectory. We denote the corresponding function of destination states from $k \in \mathcal{K}^a$ by $\mathcal{Z}^a(k)$. We model dependence on the past trajectory as lagged occurrence dependence and lagged duration dependence. With this extension we can align treated and comparison persons with respect to their unemployment experience since the beginning of the cycle. In particular, when estimating training impacts on the transition rate to employment we align treated and comparison persons in their total elapsed unemployment duration that is for the treated the prior duration in open unemployment plus the current duration in treated unemployment. Similarly, when measuring training impacts on the transition rate from employment back to unemployment we condition on the total prior unemployment duration.

Let $T_{kl} \in \mathbb{R}_+$ denote the random time spent in state k before exiting to state l , $k \in \mathcal{K}^a$, $l \in \mathcal{Z}^a(k)$ and t the elapsed duration in state k . For origin states with multiple destinations we do not observe the realizations of all exit times from state k but just that of the minimum. Let $x(t)$ denote the covariate process just before time t , $\bar{x}(t) \equiv \{x(r), 0 \leq r < t\}$ the covariate path on $[0, t)$ and X its support. We can express the conditional distribution of T_{kl} , $k \in \mathcal{K}^a$, $l \in \mathcal{Z}^a(k)$, in terms of the hazard rate:

$$\begin{aligned} \theta_{kl}(t|\tau, \bar{x}(t), v_{kl}) &\equiv \lim_{dt \rightarrow 0} \frac{\Pr[t \leq T_{kl} < t + dt | T_{kl} \geq t; \tau, \bar{x}(t), v_{kl}]}{dt} \\ &= \lambda_{kl}[t, \mathbb{1}(k \neq O)\tau, x(t)] v_{kl}, \end{aligned} \quad (3)$$

where $v_{kl} \in \mathbb{R}_+$ is an unobserved heterogeneity term while $\lambda_{kl}(\cdot): (\mathbb{R}_+ \times \mathbb{R}_+ \times X) \rightarrow \mathbb{R}_+$ denotes the part of the hazard that is a function of observed factors, i.e., elapsed time in the current state t , observed covariates $x(t)$, and lagged duration τ that enters the transition rate if origin state $k \neq O$.

Our goal is to identify the single components of the hazard rates $\theta_{kl}(\cdot)$ and the joint distribution of the unobserved heterogeneity terms, $G(\mathbf{v})$, $\mathbf{v} = \{v_{kl}: k \in \mathcal{K}^a, l \in \mathcal{Z}^a(k)\}$, which then yields the joint distribution of the durations $\{T_{kl}: k \in \mathcal{K}^a, l \in \mathcal{Z}^a(k)\}$. We proceed in a sequential way, considering first identification of the competing risks from the initial state, then of the hazard rates from the second state given the first transition and so on. In this way we identify, for a given origin state k , the joint distribution of the unobservables $\{v_{kl}: l \in \mathcal{Z}^a(k)\}$ together with the selection process into state k . Once we have identified all possible trajectories leading from open unemployment to employment, we can trace out the joint distribution of the unobservables $\{v_{kl}: k \in \mathcal{K}^a, l \in \mathcal{Z}^a(k)\}$ by varying the observed arguments of the hazard rates (i.e., $t, \tau, x(t)$) on a nonempty open set. This yields then the joint distribution of the survival times $\{T_{kl}: k \in \mathcal{K}^a, l \in \mathcal{Z}^a(k)\}$.

We describe the exact procedure and the identification results used at each stage of the unemployment-employment cycle in Section A in the Supplementary Appendix. Here, we summarize the main points. In our model with employment duration as the second outcome, our identification strategy exploits that we observe about half of the sample more than once in the initial state and that the hazard rates from treated unemployment and employment include time-varying covariates.²⁰ This allows us to avoid several restrictive assumptions, e.g., proportionality of the effects of elapsed duration, lagged duration, and observed covariates, which are commonly made in the literature on mixed hazard rate models (Abbring and van den Berg,

20. In the hazard rates from unemployment we use time-varying regressors related to the remaining entitlement period of unemployment benefits, age, and season of the year. In the hazard rates from treated unemployment, we include in addition indicators for different time intervals relative to the planned end of the program. In the employment hazards, we consider a polynomial of age (interacted with gender), indicators for season of the year and interactions between season and industry of the last job.

2003b, Proposition 3; Brinch, 2007). As suggested in equation (3), we only assume separability of the unobserved heterogeneity terms.

In our alternative model with earnings, we model the conditional distribution of earnings as a hazard rate together with the hazard rates from unemployment into employment, short-term training, and long-term training. In this model, we do not have repeated cycles per person and the hazard rates of daily earnings do not include covariates that vary with the level of earnings. Therefore, we assume that the hazard rates are separable in their components, a flexible nonnegative function of elapsed duration, $\lambda_{kl}^0(t)$, of lagged duration, $\kappa_{kl}(\mathbb{1}(k \neq O)\tau)$, of observed covariates, $\phi_{kl}(x(t))$, and unobserved heterogeneity, v_{kl} , i.e., $\theta_{kl}(t|\tau, \bar{x}(t), v_{kl}) = \lambda_{kl}^0(t)\kappa_{kl}(\mathbb{1}(k \neq O)\tau)\phi_{kl}(x(t))v_{kl}$, and that the regressors are independent of the unobserved heterogeneity terms (Abbring and van den Berg, 2003b, Proposition 2; Honoré, 1993, Theorem 3; Horny and Picchio, 2010).

3.3. *Empirical support of the conditional no-anticipation and independence assumptions*

The conditional no-anticipation assumption (1) and the conditional independence assumption (2) are important prerequisites for identification. Analogously to a static matching analysis their plausibility depends on the richness of our data and our ability to flexibly control for potentially time-varying confounders that jointly determine outcome and treatment times. We can to some extent dispense with time-constant observed covariates through including unobserved heterogeneity. The combination of three factors makes the conditional no-anticipation assumption a credible basis for causal analysis in our case: (1) our construction of the analysis sample, (2) the institutional setup determining training participation, and (3) our flexible modeling strategy based on large and informative data.

First, we construct our analysis sample in such a way that included individuals have similar recent employment histories and high incentives to comply with the instructions of the caseworker. According to the regulation of public employment services in Germany, job-seekers face tight job search requirements (Section B.1 in the Supplementary Appendix). Therefore, we focus on prime-age workers who experience a transition into unemployment after a period of continuous employment. We refrain from including individuals who start looking for a job after a period out of the labor force because they might only register at the local employment agency so long as their expected utility from receiving particular services and programs exceeds their disutility associated with tight job search requirements. As a consequence more than 90% of the persons in our sample have substantial entitlements to unemployment benefits (Section D.3 in the Supplementary Appendix). These individuals are presumably highly committed to cooperate with the local employment agency, for intrinsic reasons as well as for the fact that in case of noncompliance their generous benefits would be at risk. Overall, these sample selection criteria ensure that the individuals considered in the analysis behave in a relatively predictable way and according to the guidelines set out by public employment services.

Second, with respect to the assignment of training programs, a participation in training may take place at any point in time during the unemployment spell. Job-seekers have no entitlements regarding participation (Section B.1 in the Supplementary Appendix). However, a program assignment is compulsory for the job-seeker and noncompliance may entail benefit sanctions and the exclusion from further services. The allocation of training programs depends on the short-term supply of training slots as well as private information of the caseworker. On the one hand short notice periods and belated assignments (i.e., after the official starting date of a program) are used as a work test. On the other hand they allow caseworkers to achieve a high utilization of available training capacities and at the same time to avoid disincentives for job-seekers anticipating future program participation. As caseworkers face a very high caseload

the assignment is not targeted.²¹ As supporting evidence for this practice, we find considerable variation in the timing of training starts during the unemployment spell and modest differences in key observed characteristics between participants and nonparticipants (Sections D.2 and D.3 in the Supplementary Appendix). Anticipatory effects with regard to future training participation are therefore unlikely to occur.

The third important ingredient for conditional no-anticipation and independence is the data and the specification of the conditional hazard rates. Our administrative database is exceptionally rich with regard to the number of observations as well as the available covariate information. Large sample sizes allow us to condition on current and lagged duration, calendar time, observed covariates, and unobserved heterogeneity in a detailed and flexible way (see also Section 4.3). For instance, we use step functions to model duration dependence, which allows us to nonparametrically approximate arbitrary patterns of duration dependence. We model the impacts of training on the exit rate out of unemployment with a step function indicating different time intervals relative to the planned end of the program. In addition, we flexibly condition on calendar time at the start of the spell, current season of the year, and age of the individual. A careful specification of duration and calendar time effects is important to capture changes in the job-seeker's information set that are a function of time.

In terms of covariates, we have access to the information collected on job-seekers when they register at the local employment agency, i.e., the Supply of Applicants database. It builds the basis for the counseling activities and assignment decisions of the caseworker. It details personal characteristics, properties of the last job, and objectives of job-search. The data also include caseworker assessments of the qualifications and experience of a job-seeker and of his/her health status. Combining this with the employment register we can characterize the last employment relationship in terms of e.g. previous earnings, occupation, industry, type of position, and the reason why the job ended. In addition, we control for region of residence and local labor market conditions. We include time-varying indicators for the remaining benefit entitlement, past and current periods of sickness as well as past and current temporary suspensions of benefit payments.²²

4. IMPLEMENTATION

4.1. *Training programs analyzed*

Training schemes are the core of active labor market policy in Germany as in most advanced countries.²³ In our analysis, we focus on human capital intensive long-term training on the one hand and job search-oriented short-term training on the other hand.

Long-term training schemes comprise a variety of programs ranging from advanced vocational training and refresher courses on specific professional skills and operational techniques to comprehensive retraining in a new vocational degree within the German apprenticeship system. The former typically last between six and twelve months whereas retraining takes two to three years. Training programs may take place either in classrooms, simulated workplaces, firms, or a combination thereof. Typical examples of long-term training programs include training on

21. The average caseload in terms of registered unemployed is about 400 unemployed per caseworker (Section B.1 in the Supplementary Appendix). In addition, caseworkers counsel people who register as job-seekers but are not unemployed as well as people who do not register in the end.

22. See Section E in the Supplementary Appendix for a detailed description of the included variables.

23. See Section B.2 in the Supplementary Appendix for further information on the quantitative importance of training in the context of German active labor market policy.

marketing and sales strategies, computer-assisted bookkeeping, operating construction machines, and specialist courses in specific legal fields.

Short-term training courses last a couple of days to twelve weeks. Similar to the long-term training schemes, they may take place in classrooms or workplace-like environments. However, due to their shorter length their contents are less occupation specific and the human capital component is limited. Typical examples of short-term training schemes include job application training, basic computer courses, language courses, and short-term internships at a simulated or real workplace. Their aim is twofold. On the one hand, they provide skills that improve and facilitate job search. On the other hand, they are employed to assess a job-seeker's abilities and his/her readiness to work or to participate in a further program.

4.2. *Data*

The empirical analysis is based on an exceptionally rich administrative database, the Integrated Employment Biographies Sample (IEBS) that has recently been made available by the Research Data Center of the German Federal Employment Agency. The IEBS is a 2.2% random sample from a merged data file containing individual records out of four different administrative registers. It comprises data on employment subject to social security contributions, receipt of transfer payments during unemployment, job search, and participation in different active labor market programs. Start and end dates of the different labor market episodes are measured with daily precision.

From this data set, we extract a sample of West Germans aged twenty-five to fifty-three years who experience a transition from regular, unsubsidized employment lasting three months or longer to unemployment within the period July 1999 to December 2001. Unemployment is defined as nonemployment with at least occasional contact with the employment agency. This contact may be recorded in the data either as a benefit spell, a registered job search spell, or a period of program participation.²⁴ Unemployment spells are censored at the end date of the last contact with the employment agency if in the following three months no such contact persists. Nonemployment spells without any contact with the employment office are not considered because in that case we cannot distinguish between people looking for a job and persons out of the labor force. Also, people who are not actively looking for a job or who do not register with the employment agency are not eligible for training programs. Transitions to active labor market programs other than training are treated as independent censoring. We follow the persons in the inflow sample until the end of December 2004, and ongoing spells are censored at that date. We model the trajectory from open unemployment to employment, which may include a participation short-term training or long-term training. In addition, we consider the transition from employment back to unemployment. Individuals in the sample may have multiple unemployment and employment spells if they experience multiple transitions between unemployment and employment.

In the alternative model with earnings, we model the pathway from open unemployment to employment and the average earnings per calendar day within the first 365 days upon reemployment. Days during which a person is not employed contribute with a zero. Earnings are censored at the social security threshold and if the time between the end of unemployment and the end of our observation period is less than 365 days. In the model with earnings, we use only the first cycle for every individual because we cannot measure earnings dynamics at the same frequency as employment dynamics. Total earnings are reported only once every calendar

24. For 12 out of 13,859 training spells, the training spell constitutes the first contact with the employment agency recorded in the data.

TABLE 1
Unemployment and employment outcomes by treatment status

	Unemployment		Employment		Earnings	
	$E(T T \leq 1825)$	Median	$E(T T \leq 1825)$	Median	$E(T T \leq 144)$	Median
No training	219	160	375	284	49.87	49.25
Short training	235	172	361	304	45.08	43.29
Long training	570	607	477	506	50.16	50.13
All	261	167	380	287	49.56	48.87
Observations	87,250		56,758		30,190	

Unemployment and employment durations are measured in days, daily earnings in € (real values, reference year 2000).

year in employment relationships that last longer than a year. We model the conditional density of daily earnings in the same way as the transition rates into different labor market states, i.e., using a mixed proportional hazard rate specification.

Figure 1 above illustrates the different transitions of the 45,459 people in our sample. In total, there are 87,250 unemployment spells, of which 8,279 (5,580) lead to a participation in short-term training (long-term training), and there are 56,758 employment spells. About half of the individuals in the sample experience multiple unemployment spells and about 30% have more than one employment spell. In contrast, only 8% (2%) of the treated with short-term training (long-term training) participate more than once in the same program.

Table 1 provides summary statistics on the distributions of unemployment and employment durations as well as wages, by treatment status and for the entire sample. The means in Table 1 refer to the expected value conditional on the truncated distribution of T until truncation point $\bar{t} > 0$.²⁵ It is calculated as follows:

$$E[T|T \leq \bar{t}] = \frac{\int_0^{\bar{t}} r dF(r)}{F(\bar{t})} = \frac{\int_0^{\bar{t}} S(r) dr - \bar{t}S(\bar{t})}{1 - S(\bar{t})}, \quad (4)$$

where $F(t)$ is the cumulative distribution function and $S(t)$ the survivor function. We estimate $S(t)$ with the Kaplan and Meier (1958) estimator. In sum, the evidence in Table 1 suggests that long-term training has beneficial effects on employment duration but at the same time seems to strongly increase unemployment duration. Short-term training, in contrast, tends to reduce unemployment duration but seems to have mixed effects on employment and earnings. However, it is important to stress that these patterns reflect a mixture of causal and selection effects. It remains to be seen to what extent they persist after taking into account the dynamic selection into treatments and into outcome states.²⁶

4.3. Model specification

We adopt a piecewise constant exponential model for the hazard rates. For the index functions, we use flexible, linear in parameters specifications to model the dependence on observed covariates $x(t)$ and lagged duration τ . We use piecewise constant specifications to model the dependence on

25. We choose the truncation points for unemployment and employment duration according to the limits of our observation period and for daily earnings according to the social security threshold in 2000.

26. Sections C, D, and E in the Supplementary Appendix provide further details on the administrative data base, additional descriptive evidence and a complete list of the covariates used in the estimations.

elapsed duration t and time dependence of training impacts during unemployment. Specifically, in our model with employment, the hazard rate for the transition from state k to l is given by:

$$\begin{aligned}\theta_{kl}(t|\tau, x(t), v_{kl}) &= \lambda_{kl}[t, \mathbb{1}(k \neq O)\tau, x(t)] v_{kl} \\ &= \exp[h_{kl}(t, \mathbb{1}(k \neq O)\tau, x(t))\beta_{kl}] v_{kl},\end{aligned}$$

where $h_{kl}(\cdot)$ is a function of the observed components elapsed time t , lagged duration τ (that enters the hazard rate if state $k \neq O$) and observed covariates $x(t)$, and β_{kl} is a conformable coefficient vector. In our alternative model with earnings, the hazard rate for the transition from state k to l is modeled as $\theta_{kl}(t|\tau, x(t), v_{kl}) = \exp[b_{kl}(t)\gamma_{kl} + d_{kl}(\mathbb{1}(k \neq O)\tau)\delta_{kl} + f_{kl}(x(t))\zeta_{kl}] v_{kl}$, with $b(\cdot)$, $d(\cdot)$, $f(\cdot)$ some flexible functions and γ , δ , ζ conformable coefficient vectors.

We model the joint distribution of the unobservables, $\{v_{kl} : k \in \mathcal{K}^a, l \in \mathcal{Z}^a(k)\}$, with a discrete masspoint distribution that in principle allows to approximate any arbitrary discrete or continuous distribution (Heckman and Singer, 1984). In particular, we adopt a factor loading specification with two independent underlying factors, w_1 and w_2 :

$$v_{kl} = \exp(\alpha_{kl1} w_1 + \alpha_{kl2} w_2), \quad k \in \mathcal{K}^a, l \in \mathcal{Z}^a(k),$$

where α_{kl1} and α_{kl2} denote the factor loadings on the fundamental unobserved factors w_1 and w_2 , respectively, for a transition from state k to l . We normalize w_1 and w_2 to have support $\{-1, 1\}$ and in addition constrain one of the α_{kl2} to equal zero.²⁷ The two-dimensional factor loading model imposes no restrictions on the covariance matrix of the unobserved heterogeneity terms. Let $\mathbf{w} = (w_1, w_2)'$ and A be the matrix of factor loadings with rows $A_{kl} = (\alpha_{kl1}, \alpha_{kl2})$, $k \in \mathcal{K}^a, l \in \mathcal{Z}^a(k)$. Then the variance-covariance matrix of the unobserved heterogeneity terms is given by $\text{Var}(\ln(\mathbf{v})) = A \text{Var}(\mathbf{w}) A'$.

In our preferred specifications, each of the two factors has two masspoints. We experimented with different numbers of masspoints for the latent factors. The maximum number that led to a significant improvement of the log likelihood function was three masspoints for each factor in the model with employment duration. However, the model fit in terms of a comparison between the original and simulated duration distributions differed only marginally from that based on two masspoints for each factor. Importantly, the model with two masspoints for each factor is computationally more attractive. Since we normalize the locations of the masspoints we can avoid evaluation of their derivatives. This speeds up computation by a factor of two to four, which is important when it comes to bootstrapping standard errors.

4.4. Estimation

Conditional on the observed covariates and the unobserved determinants, the joint density of the four durations for individual i is given by:

$$f_i = \prod_{j=1}^{N_i} \prod_{p=1}^{P_{ij}} \theta_p(t_{ijp} | \tau_{ijp}, x_{ij}(t_{ijp}), v_{ip})^{c_{ijp}} \exp\left[-\int_0^{t_{ijp}} \theta_p(r | \tau_{ijp}, x_{ij}(r), v_{ip}) dr\right],$$

where N_i is the number of cycles of person i , P_{ij} is the total number of possible origin destination state pairs associated with person i 's j -th cycle, and c_{ijp} is a censoring indicator that equals one

27. All hazards contain an intercept.

if the observation period of person i in cycle j ends with a transition from the origin state to the destination state indexed by p . $\theta_p(\cdot)$ is the hazard rate that depends on elapsed duration in the current state t , lagged duration τ , observed covariates $x(t)$, and unobserved heterogeneity v . Since we allow for nonzero correlations of the unobserved heterogeneity terms in the different hazard rates, the likelihood function is not separable by individual and spell type. Thus, the individual likelihood contribution conditional on observed covariates and integrated over the vector of unobserved heterogeneity terms, \mathbf{v} , is:

$$L_i = \int \cdots \int \prod_{j=1}^{N_i} \prod_{p=1}^{P_{ij}} \theta_p(t_{ijp} | \tau_{ijp}, x_{ij}(t_{ijp}), v_{ip})^{c_{ijp}} \times \\ \exp \left[- \int_0^{t_{ijp}} \theta_p(r | \tau_{ijp}, x_{ij}(r), v_{ip}) dr \right] dG(\mathbf{v}).$$

The final models involve the estimation of more than 350 and 500 parameters, respectively. The maximization routine uses a Newton-Raphson algorithm with analytic first and second derivatives.²⁸ To obtain standard errors, we implement the semi-parametric bootstrap procedure suggested in Gaure, Røed, van den Berg and Zhang (2010). For the coefficients on the observed covariates we draw values from their joint normal distribution.²⁹ Based on these draws we estimate a constrained model keeping the coefficients on the observed covariates fixed and maximizing over the coefficients associated with the unobserved heterogeneity distribution. We perform 250 bootstrap replications.³⁰ The full estimation results and the simulated model fit are displayed in Sections F and G in the Supplementary Appendix.

5. CAUSAL ESTIMATES OF TRAINING IMPACTS

5.1. *Impacts of training on the treated*

We use the parameter estimates to simulate the marginal distributions of treatment and nontreatment outcomes in the sample of treated persons. In particular, we consider all persons who receive training within the first 731 days of their first unemployment spell.³¹ We exclude later training starts in order to avoid extrapolations beyond the time horizon of our data. We simulate ten unemployment-employment cycles for each person and treatment status.³² From the simulated durations, we estimate the marginal distributions of treatment and nontreatment outcomes with the Kaplan and Meier (1958) estimator. We further assess training impacts through contrasting summary statistics of the marginal distributions, e.g. the truncated mean given in equation (4). We compute standard errors for the summary statistics by rerunning the simulation procedure 250 times using the parameter values from the bootstrap replications.³³

28. The maximum likelihood evaluator is implemented in Stata MP Version 11.1 and its matrix language Mata.

29. Van der Vaart (1996) shows that, in the exponential hazard model with nonparametric unobserved heterogeneity, the coefficients attached to observed covariates are asymptotically normal.

30. We use the bootstrap distribution of the unobserved heterogeneity parameters to test them for normality. The tests suggest no deviation from normality for 16 out of 17 parameters in model 1 (employment) and 7–9 parameters out of 13 in model 2 (earnings as second outcome).

31. We obtain a subsample 7,046 of treated persons, 3,489 of which participate in short-term training, and 3,557 in long-term training.

32. Further details are provided in Section H in the Supplementary Appendix.

33. As the summary statistics on training impacts are smooth functionals of the survivor function we treat them as normal distributed when conducting hypothesis tests.

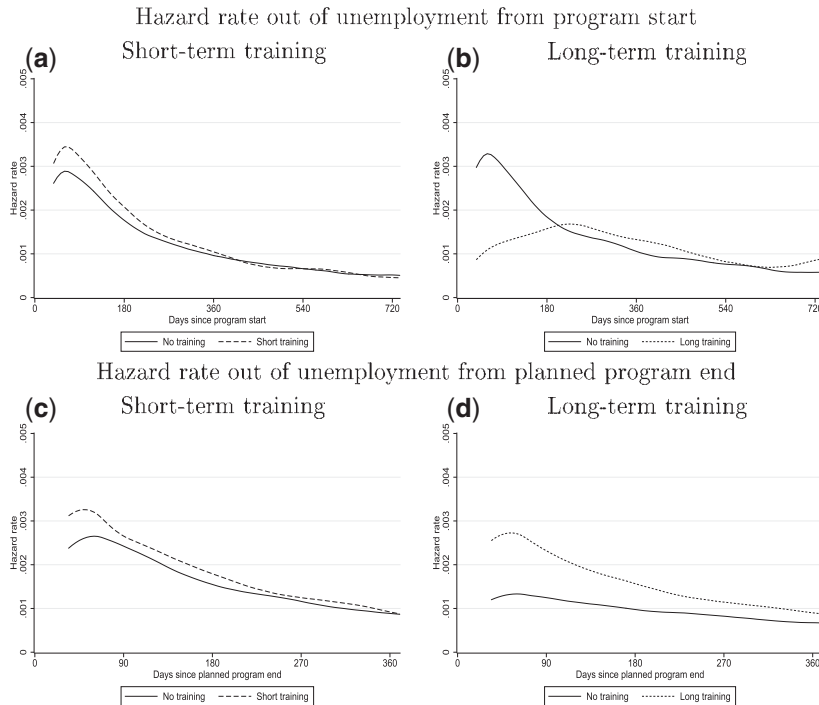


FIGURE 2

Impacts of training on the hazard rate out of unemployment.

Notes: Kaplan-Meier (1958) estimates of the hazard rates are based on simulated unemployment durations with and without training for a subset of the originally treated persons. The bandwidth for the kernel smooth of the hazard rates is 30 days. Estimates within 30 days of the left boundary are omitted.

Our data includes information on the planned and actually realized end dates of training programs. The planned enrolment length is assigned at the start of the program and may be shorter or longer than the realized duration.³⁴ We model duration dependence of treatment effects with dummy variables indicating different time intervals relative to the planned program end. This allows us to contrast in-program effects with postprogram effects of training. Thus, we consider two measures for the impact of training on unemployment: the effect on the remaining time counted from the start of training and from the planned end of training. Figure 2a and b depict the simulated hazard rates with and without training from the program start onwards, figure 2c and d show the hazard rates starting from the planned program end.

Figure 2a shows that participants in short-term training exit at a faster rate, compared with the situation without training, already from the beginning of the program onwards. Thus, a participation in short-term training, that has a median planned duration of four weeks, has no noticeable lock-in effect. To the contrary, it immediately helps participants in getting reemployed. The effect reaches its maximum about 65 days after the program start. At that date, the hazard rate with short-term training is 18% higher than that without training. However, it then fades away relatively quickly. According to figure 2c, the difference between the hazard rates with and without training is already very small three months after the planned program end.

34. Waller (2009) reports that about 20% of the persons who attend a training program for more than one week drop out early, i.e., before completing 80% of the scheduled enrolment length.

TABLE 2
Nontreatment outcomes (NTO) and treatment effects on the treated (TT) for remaining unemployment duration

Remaining time in unemployment from program start					
	$E(T T \leq 731)$	$Pr(T > 731)$	P15	P30	P45
Short-term training					
NTO	191.39 (2.20)	0.39 (0.01)	54 (1.44)	122 (4.01)	264 (11.39)
TT	-16.37 (4.02)***	-0.04 (0.02)*	-7 (2.83)**	-21 (6.99)***	-59 (21.02)***
Long-term training					
NTO	187.68 (1.81)	0.35 (0.01)	46 (1.24)	106 (2.92)	225 (8.52)
TT	95.01 (4.19)***	0.08 (0.03)***	97 (8.61)***	154 (15.19)***	211 (37.66)***
Remaining time in unemployment from planned program end					
	$E(T T \leq 365)$	$Pr(T > 365)$	P15	P30	P45
Short-term training					
NTO	121.54 (0.84)	0.53 (0.01)	55 (1.52)	135 (4.65)	320 (14.00)
TT	-10.84 (1.57)***	-0.06 (0.02)***	-23 (3.12)***	-35 (8.81)***	-94 (24.89)***
Long-term training					
NTO	143.64 (1.41)	0.69 (0.01)	118 (4.29)	332 (13.37)	—
TT	-24.54 (2.97)***	-0.17 (0.03)***	-69 (6.13)***	-201 (19.35)***	—

Unemployment durations are measured in days. P15, P30, P45 refer to the percentiles of the cumulative distribution function. Calculations are based on simulated unemployment durations with and without training for a subset of the originally treated persons. Bootstrapped standard errors based on 250 replications in parentheses. *, **, and *** denote significance at the 10, 5 and 1% level, respectively.

The picture is rather different for long-term training programs that have a median planned duration of 201 days. In figure 2b, we see that, during participation, people exit unemployment at a much lower rate compared with the situation of nonparticipation. Only after about more than 200 days since program start, when the majority has completed training, the hazard rate recovers to a level slightly exceeding the one without participation. However, once we concentrate on the period following the scheduled program end, in figure 2d, we see that long-term training has strong and persistent positive effects on the exit rate to employment. During the first three months after the planned program end, the hazard rate out of unemployment is about twice as high for the participants compared with the situation had they not participated. This effect slowly declines over time. One year after the planned program end, the hazard rate with participation in long-term training is about 40% higher than without training.

Table 2 provides summary statistics and bootstrapped standard errors on the impacts of short- and long-term training on the remaining time in unemployment. In each table, the row labeled “NTO” refers to the simulated nontreatment outcome and the row labeled “TT” to the simulated treatment effect on the treated. The columns labeled “ $E[T|T \leq 731]$ ” and “ $E[T|T \leq 365]$ ” refer to the truncated means that are calculated according to equation (4), now for the remaining time in unemployment until 731 days after the start of training and 365 days after the planned end, respectively. Counting from the program start, we see that short-term training reduces the expected remaining time in unemployment by a statistically significant sixteen days, about 8.6%, whereas participating in long-term training significantly increases it by ninety-five days or 50.6%. Considering the time from the planned program end onwards, we see that long-term training has a stronger beneficial effect than short-term training: the latter reduces the average remaining duration in unemployment by eleven days (9.0%), whereas the reduction achieved with long-term training is twenty-five days (17.4%).

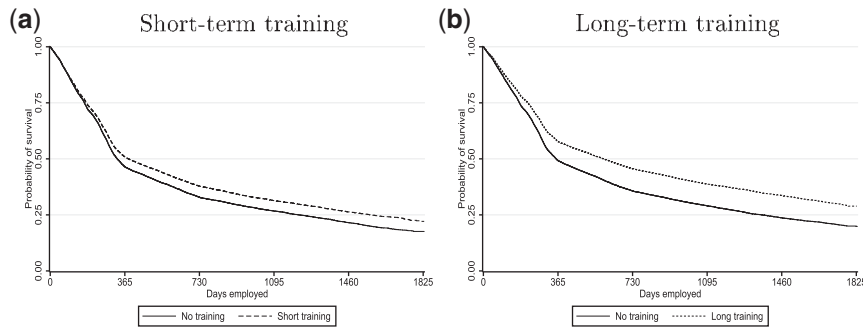


FIGURE 3

Impacts of training on the probability of survival in employment,

Notes: Kaplan-Meier (1958) estimates of the probability to remain employed are based on simulated employment durations with and without training for a subset of the originally treated persons.

Figure 3 displays the simulated impacts of training on the probability to stay employed beyond a given elapsed duration. For both training programs the probability to remain employed is higher with training than without at longer elapsed durations. The vertical difference between the survivors with and without training increases up to about 1.5 years after the beginning of the employment spell and remains constant thereafter. When the vertical difference stabilizes, the effect of long-term training is about twice as large as that of short-term training. Thus, long-term training increases the stability of employment much more strongly than short-term training. Taking the horizontal difference between the survivor functions with and without training in each graph we obtain the quantile treatment effect. Figure 3 shows that the quantile treatment effects are increasing with decreasing percentiles of the survivor functions. The substantial heterogeneity of the quantile treatment effects suggests that there is also a considerable heterogeneity in the distribution of individual treatment effects.³⁵

Table 3 translates the rather qualitative findings obtained from figure 3 into summary statistics and provides bootstrapped standard errors. The impact of short-term training on the expected employment duration (truncation point is 1,825 days) is seven days, 1.4%, and insignificant. The effect of long-term training on the mean employment duration is with twenty-three days or 5% somewhat bigger but still insignificant. However, at the median and the 70th percentile the horizontal distance between the cumulative distribution functions with and without long-term training is much larger and increasing. Similarly, the probability of survival in employment five years after the start of the spell is 9 percentage points or 45% higher with long-term training compared with the situation without. The corresponding number for short-term training is 5 percentage points, a bit less than 30%. Twenty-nine percent of the survival times in employment associated with long-term training exceed 1,825 days compared with 22% of those with short-term training. This suggests, that the truncated mean underestimates in particular the effect of long-term training on the subsequent employment duration.

Table 4 shows the impacts of training on average daily earnings during the first year after reemployment. Short-term training has no effect on earnings. The point estimate is almost zero and not significant statistically. Contrary to the descriptive findings, long-term training leads to significant earnings gains of €3.37 or 7% a day. These estimates likely underestimate the full

35. See Heckman, Smith, and Clements (1997) for a discussion of properties of the distribution of individual treatment effects that can be inferred from the marginal distributions of treatment and nontreatment outcomes.

TABLE 3
Nontreatment outcomes (NTO) and treatment effects on the treated (TT) for employment duration

	$E(T T \leq 1825)$	$Pr(T > 1825)$	P30	P50	P70
Short-term training					
NTO	435.57 (9.07)	0.17 (0.01)	207 (4.72)	330 (7.20)	869 (38.67)
TT	6.99 (14.31)	0.05 (0.02)***	11 (11.22)	48 (40.20)	279 (129.57)**
Long-term training					
NTO	445.96 (7.17)	0.20 (0.01)	216 (4.03)	354 (8.50)	1050 (47.70)
TT	22.71 (16.61)	0.09 (0.03)***	35 (15.20)**	215 (92.31)**	682 (176.65)***

Employment duration is measured in days. P30, P50, P70 refer to the percentiles of the cumulative distribution function. Calculations are based on simulated employment durations with and without training for a subset of the originally treated persons. Bootstrapped standard errors based on 250 replications in parentheses. *, **, and *** denote significance at the 10, 5, and 1% level, respectively.

TABLE 4
Nontreatment outcomes (NTO) and treatment effects on the treated (TT) for daily earnings

	$E(T T \leq 144)$	$Pr(T > 144)$	P25	P50	P75
Short-term training					
NTO	46.29 (0.83)	0.015 (0.001)	26.22 (0.66)	45.54 (1.06)	63.55 (1.03)
TT	-0.05 (0.58)	-0.003 (0.001)***	0.20 (0.45)	-0.06 (0.72)	-0.62 (0.70)
Long-term training					
NTO	47.02 (0.88)	0.014 (0.001)	26.58 (0.70)	46.67 (1.12)	64.67 (1.12)
TT	3.37 (1.48)**	0.001 (0.003)	3.59 (1.33)***	4.58 (1.69)***	3.56 (2.05)*

Earnings per calendar day are measured in € (real values, reference year 2000). P25, P50, P75 refer to the percentiles of the cumulative distribution function. Calculations are based on simulated earnings with and without training for a subset of the originally treated persons. Bootstrapped standard errors based on 250 replications in parentheses. *, **, and *** denote significance at the 10, 5 and 1% level, respectively.

monetary return to training as our earnings measure refers to the first year after reemployment only. In fact, our findings for employment stability suggest that especially persons treated with long-term training accumulate more work experience than without training over the longer run. This may lead to additional earnings gains in the medium and long run.³⁶

5.2. Importance of the timing of training

We model the time dependence of treatment effects as proportional shifts of the hazard rate out of unemployment that vary with time relative to the planned program end. In addition, we include a linear and a quadratic term of the log unemployment duration at program start in the hazard rate. This specification allows for complex effects of the timing of training during the unemployment spell. In order to assess the *ceteris paribus* effect of changing the starting date of training during unemployment, we predict the truncated mean unemployment duration associated with different starting dates for a reference person with characteristics at the mean or mode of each covariate. Specifically, this person is a male German aged thirty-eight years

36. To get an idea of the persistence of training impacts in later unemployment and employment spells, we include indicators for lagged training participation. Overall, the estimates indicate some persistence across spells. However, positive effects on employment stability tend to be offset by negative effects on unemployment duration.

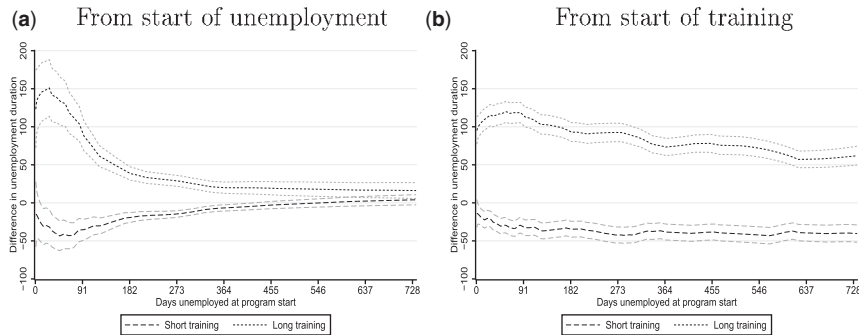


FIGURE 4

Training impacts on unemployment duration by starting date.

Notes: Calculations are based on the predicted means measured from the start of the spell (a) and from the start of training (b), respectively, for a reference person, cf. Section 5.2. The truncated means are calculated according to equation (4). The truncation point is 1,825 days in the left panel and 731 days in the right. The gray lines are 95% confidence intervals based on bootstrapped standard errors with 250 replications.

at the start of the spell, married without children, living in the German federal state of North Rhine-Westphalia, holding a secondary schooling degree reached at the end of the 10th grade (Realschule) and a vocational training degree, previously employed as a bluecollar craftsman in the trade and transport sector, with log daily earnings of 3.9 (i.e. in the third quartile), entitled to 340 days of unemployment benefits, considered as having relevant vocational qualification and experience by the caseworker, and starting his unemployment spell in the first quarter of 2000.³⁷ The planned program durations are set to their medians at twenty-six (short-term training) and 201 days (long-term training), respectively. The unobserved heterogeneity terms are set to their mean values.

Figure 4a shows the impact of the timing of training on the truncated expected unemployment duration measured from the start of unemployment until Day 1,825 and the right panel shows the impact on the truncated expected remaining unemployment duration measured from the start of training until Day 731 after training start.³⁸ It suggests that, in absolute value, the impact of both short-term and long-term training on the total unemployment duration is strongest when training is started early during unemployment, but with opposite signs. A participation in long-term training starting during the first three months of unemployment increases the total expected time in unemployment by up to 151 days, whereas a participation in short-term training in the same period decreases the total expected time in unemployment by up to 43 days. With increasing elapsed unemployment duration at program start the impacts of both training programs on the total expected unemployment duration decrease in absolute value. The effect of long-term training converges to +16 days, whereas that of short-term training approaches to zero as the time in unemployment at program start increases.

Figure 4b displays the impact of the timing of training on the remaining time in unemployment counted from the start of the program. The impact of short-term training is largely constant at

37. The original estimation includes time-varying dummies indicating different seasons of the year. In the simulation, we assign each of these variables a time-constant value representing their share in a given calendar year. This way we obtain expected outcome durations that are seasonally adjusted.

38. The truncated expectations are calculated according to equation (4).

about –40 days regardless of the starting date during unemployment. In contrast, the impact of long-term training on the remaining time in unemployment displays a declining pattern, from about +115 days for program starts during the first three months to +60 days for program starts in the eighth quarter after the beginning of unemployment. The different patterns in figure 4a and b reflect that longer survival times are weighted differently in the two graphs. While the expected value from the start of training (figure 4b) conditions on survival in unemployment until program start, the expected value from the start of unemployment (figure 4a) accounts for the possibility that the unemployment spell may terminate before the start of training. Therefore, in figure 4a, training impacts at later points in time during unemployment receive a lower weight than training impacts occurring at the start of unemployment. This difference in weighting largely drives the apparently diminishing effectiveness of short-term training in figure 4a, while the analog does not hold for long-term training. In fact, our findings suggest that long-term training programs starting later during unemployment lead to a considerably smaller increase in unemployment duration.

5.3. *Effect heterogeneity across observed characteristics*

Our estimated models flexibly account for heterogeneous treatment effects across observed and unobserved variables. To examine the differences across observed characteristics, we aggregate the simulated treatment and nontreatment outcomes in a given subgroup, e.g., men, and calculate the treatment effect in this subgroup. The entries in tables 5 and 6 show the program impact on the treated for the subgroup given in the row label. The treatment effects refer to the truncated means for the outcomes remaining time in unemployment from program start and planned program end as well as daily earnings. The truncation points are 731 days, 365 days, and €144, respectively. For the outcome employment duration, we present results on the quantile treatment effect evaluated at the median and on the probability to remain employed after 1,825 days. The last row for each set of subgroups provides the p -value of a Wald test that the treatment effects are equal across subgroups.³⁹

Several interesting differences across groups emerge. Female participants in long-term training tend to benefit more than males, whereas no consistent gender differences exist for short-term training. Long-term training increases the remaining time in unemployment measured from the program start by 87 days (44.6%) for women and 102 days (56.8%) for men. Female participants in long-term training achieve higher employment gains of 291 days (64.8%) at the median and earnings gains of €4.37 (10.4%) on average per calendar day as opposed to 115 days (36.7%) and €2.69 (5.2%), respectively, for males. For older workers aged forty-five years and above, short-term training tends to work less well than for the younger age groups. However, the null hypothesis of equal of treatment effects across age groups cannot be rejected for any of the impact measures at conventional significance levels. Comparing the results for foreign nationals to those of Germans participating in long-term training, we find very small effects on employment duration for foreign people but large effects for Germans (e.g., a quantile treatment effect of +312 days, 84.7%). However, foreign nationals participating in long-term training exhibit higher earnings gains of €4.83 (11.1%) on average per calendar day as opposed to €2.80 (5.8%).

High-skilled participants in long-term training holding a university degree or previously working as senior officials, managers and professionals exhibit very strong lock-in effects, i.e., a 107–140 days (around 60%) increase in the remaining unemployment duration measured from

39. As an alternative, in Section I in the Supplementary Appendix, we provide evidence from hypothesis tests that indicate whether the proportionate treatment effect on the hazard rate is different for a particular subgroup keeping everything else constant.

TABLE 5
Heterogeneous treatment effects of short-term training

	Remaining days in unemployment from		Employment		Earnings
	program start E[T T ≤ 731]	planned prog. end E[T T ≤ 365]	P50	Pr(T > 1825)	E[T T ≤ 144]
Gender					
Male	-20.19 (4.56)	-11.43 (1.88)	30 (23.80)	0.053 (0.021)	-0.91 (0.68)
Female	-11.58 (5.28)	-10.13 (1.71)	92 (56.97)	0.043 (0.020)	1.24 (0.78)
<i>p</i> -value of Wald test of equality	0.123	0.474	0.140	0.588	0.014
Nationality					
German	-15.22 (4.20)	-10.39 (1.70)	61 (51.82)	0.042 (0.019)	-0.72 (0.61)
Foreign	-18.94 (4.36)	-11.81 (1.84)	52 (30.38)	0.062 (0.022)	1.49 (0.92)
<i>p</i> -value of Wald test of equality	0.190	0.387	0.800	0.276	0.018
Age					
25–34 (years)	-17.72 (4.38)	-12.13 (1.89)	36 (30.83)	0.039 (0.019)	0.04 (0.60)
35–44 (years)	-17.55 (4.47)	-10.51 (2.00)	64 (47.29)	0.052 (0.018)	-0.12 (0.58)
≥45 (years)	-10.87 (4.65)	-8.57 (1.94)	64 (45.28)	0.064 (0.027)	-0.08 (0.65)
<i>p</i> -value of Wald test of equality	0.104	0.202	0.301	0.424	0.600
Education					
No voc. degree and missing	-15.23 (4.12)	-11.57 (1.69)	17 (14.15)	0.014 (0.017)	0.52 (0.61)
Vocational training degree	-18.05 (4.37)	-11.24 (1.73)	102 (54.29)	0.072 (0.021)	-0.17 (0.61)
University degree	-7.81 (5.48)	-3.49 (3.35)	104 (57.74)	0.011 (0.028)	-0.16 (0.74)
<i>p</i> -value of Wald test of equality	0.099	0.025	0.115	0.000	0.088
Occupation					
Elementary occupations	-12.72 (4.10)	-9.45 (2.37)	19 (17.18)	0.065 (0.022)	0.46 (0.65)
Agricultural occ. and missing	15.12 (10.85)	0.67 (5.56)	10 (17.74)	-0.026 (0.052)	0.12 (0.83)
Plant/machine operator	-16.51 (4.24)	-13.42 (2.84)	31 (26.07)	0.018 (0.030)	0.10 (0.72)
Craft and related occupations	-23.85 (5.32)	-13.84 (2.11)	16 (11.35)	0.045 (0.022)	-0.46 (0.73)
Service worker	-14.47 (5.26)	-9.03 (2.71)	73 (49.19)	0.054 (0.027)	1.53 (0.72)
Clerk	-17.11 (5.55)	-10.21 (2.09)	99 (64.55)	0.065 (0.026)	-0.38 (0.96)
Techn., assoc. professional	-15.92 (5.79)	-16.57 (2.41)	129 (63.75)	0.077 (0.032)	-0.19 (0.86)
Professional	-10.99 (6.69)	-1.55 (3.17)	44 (63.59)	0.029 (0.025)	-0.38 (0.79)
Senior official, manager	-29.48 (10.02)	-8.85 (4.62)	105 (78.81)	0.028 (0.051)	0.62 (1.09)
<i>p</i> -value of Wald test of equality	0.014	0.000	0.008	0.324	0.034

Unemployment and employment durations are measured in days, earnings per calendar day in € (real values, reference year 2000). P50 refers to the difference between treatment and nontreatment outcomes at the median of the cumulative distribution functions. Calculations are based on simulated outcomes with and without training for a subset of the originally treated persons. Bootstrapped standard errors based on 250 replications in parentheses.

TABLE 6
Heterogeneous treatment effects of long-term training

	Remaining days in unemployment from program start $E[T T \leq 731]$	planned prog. end $E[T T \leq 365]$	Employment		Earnings $E[T T \leq 144]$
			P50	$Pr(T > 1825)$	
Gender					
Male	102.22 (4.27)	-24.19 (3.32)	115 (74.82)	0.077 (0.030)	2.69 (1.57)
Female	87.31 (4.53)	-24.89 (3.27)	291 (133.84)	0.110 (0.035)	4.37 (1.53)
<i>p</i> -Value of Wald test of equality	0.000	0.805	0.026	0.098	0.051
Nationality					
German	95.33 (4.19)	-26.20 (3.03)	312 (118.11)	0.118 (0.032)	2.80 (1.51)
Foreign	94.52 (5.26)	-20.64 (3.69)	46 (58.59)	0.029 (0.035)	4.83 (1.65)
<i>p</i> -Value of Wald test of equality	0.829	0.048	0.000	0.000	0.048
Age					
25–34 (years)	105.07 (4.21)	-21.47 (3.66)	183 (91.28)	0.087 (0.032)	3.11 (1.47)
Age 35–44 (years)	84.83 (4.64)	-28.68 (3.25)	236 (96.94)	0.088 (0.032)	3.57 (1.48)
≥ 45 (years)	93.66 (5.31)	-21.00 (4.07)	228 (88.52)	0.113 (0.034)	3.51 (1.53)
<i>p</i> -value of Wald test of equality	0.000	0.018	0.006	0.346	0.059
Education					
No voc. degree and missing	93.04 (4.45)	-22.45 (3.57)	49 (52.05)	0.043 (0.031)	4.12 (1.47)
Vocational training degree	92.77 (4.42)	-25.03 (3.08)	311 (130.21)	0.120 (0.034)	3.32 (1.56)
University degree	125.15 (5.84)	-34.09 (6.48)	350 (213.10)	0.099 (0.045)	1.85 (1.35)
<i>p</i> -value of Wald test of equality	0.000	0.129	0.005	0.000	0.005
Occupation					
Elementary occupations	93.46 (5.37)	-20.58 (4.59)	60 (54.53)	0.022 (0.034)	5.53 (1.50)
Agricultural occ. and missing	97.80 (8.24)	-45.47 (13.15)	47 (56.98)	0.077 (0.039)	4.09 (1.62)
Plant/machine operator	89.99 (4.67)	-23.17 (5.01)	54 (56.86)	0.078 (0.031)	5.06 (1.64)
Craft and related occupations	90.58 (4.61)	-30.85 (4.02)	64 (59.14)	0.087 (0.036)	3.20 (1.93)
Service worker	99.54 (6.57)	-11.59 (5.33)	161 (90.64)	0.073 (0.033)	6.80 (1.55)
Clerk	89.61 (4.87)	-28.53 (4.06)	354 (151.83)	0.111 (0.039)	1.63 (1.69)
Techn., assoc. professional	99.84 (4.28)	-22.16 (4.19)	432 (170.25)	0.166 (0.038)	1.22 (1.66)
Professional	106.97 (5.72)	-29.37 (5.88)	207 (167.75)	0.091 (0.040)	0.33 (1.43)
Senior official, manager	139.72 (11.63)	-33.95 (13.24)	427 (248.20)	0.166 (0.155)	8.60 (2.04)
<i>p</i> -value of Wald test of equality	0.000	0.007	0.075	0.000	0.000

Unemployment and employment durations are measured in days, earnings per calendar day in € (real values, reference year 2000). P50 refers to the difference between treatment and nontreatment outcomes at the median of the cumulative distribution functions. Calculations are based on simulated outcomes with and without training for a subset of the originally treated persons. Bootstrapped standard errors based on 250 replications in parentheses.

the program start. For participants in long-term training previously working in low- to medium-skilled occupations the corresponding figure lies between +90 and +100 days (around 50%). Persons with low formal qualifications as well as persons previously working in elementary occupations exhibit small employment gains: at the median the effect is +49 days (16.2%) for low skilled and +60 days (20.1%) for elementary occupations, respectively. In contrast, people with medium formal qualifications and people previously working in medium-skilled job that are intensive in analytic and interactive tasks show very strong employment gains. Technicians and associate professionals, for instance, gain 432 employment days (89.8%) at the median and have a 16.6 percentage points (69.3%) higher probability to remain employed after five years. High skilled and persons in high-skilled occupations also show big point estimates for employment duration, but the standard errors tend to be large as well.

When it comes to earnings impacts of long-term training the ordering tends to be reversed. Persons with low formal education and persons previously working in low- and medium-skilled manual jobs (in particular, elementary occupations and service workers) reap the highest earnings gains of €4.12, €5.53, and €6.80, respectively. In relative terms, these earnings gains are in the order of 10.3, 15.1, and 18.9%, respectively, a substantial amount from an economic point of view. Senior officials and managers, experience an earnings gain of €8.60 per calendar day, which corresponds to a gain of 11.9% in relative terms. The earnings gains for persons holding a university degree as well as for people previously working in high- and medium-skilled occupations that are intensive in analytic and interactive tasks, in contrast, are small (between €0.33 and €1.85).

The differing impacts of long-term training across occupation groups may also contribute to explaining the differences across gender. In our sample, 36.8% of female job-seekers previously worked as clerk or technician and associate professional as opposed to only 18.1% of the males. These occupations reap the highest employment gains. 23.9% of the women as opposed to 7.5% of the men previously worked as service workers, a group with above average earnings gains. Forty-nine percent of the male unemployed previously worked in production occupations (plant/machine operator or craft and related occupations) that show below average employment gains from long-term training. However, the earnings gains for these occupations tend to be above average. Similarly, the smaller employment and higher earnings effects for foreign nationals as opposed to Germans participating in long-term training may be associated with their higher shares in elementary and production occupations.

As regards short-term training, the group of technicians and associate professionals exhibits exceptionally high benefits in terms of employment stability with a quantile treatment effect of +129 days (25.5%). For the other occupation groups the quantile treatment effects for employment duration are smaller. However, we find sizeable impacts of 5–8 percentage points on the probability to be still employed after five years for five out of nine occupation groups. Overall, the differences across skill and occupation groups do not suggest a coherent interpretation.

In sum, these patterns suggest that long-term training programs facilitate the occupational advancement of heterogeneous groups of participants. However, their beneficial effects manifest in different ways across different groups. While they have large positive effects on employment stability for people in medium- to high-skilled analytic and interactive jobs, there are almost no employment effects for people in low- to medium-skilled service and production occupations. However, the latter groups show above average earnings gains, whereas the former earn hardly more than without training. The often sizeable employment and earnings gains come at the cost of prolonged unemployment spells. The opportunity cost of participating in long-term training is highest for people holding a university degree as well as for senior officials and managers. Short-term training programs do not harm any of the groups considered but their beneficial effects are generally limited in magnitude.

5.4. *Cost-benefit assessment*

From a policy point of view it is important to know whether the gains of participating in short-term or long-term training outweigh the costs. In order to get an idea of the cost-effectiveness of the programs consider the following back-of-the-envelope calculations. A short-term training course costs on average €560 and a long-term training course €5,850 (cf. Table B1 in the Supplementary Appendix). The employment agency pays on average €1,050 per month for an unemployed entitled to unemployment benefits. Extrapolating beyond the limits imposed by our data we can calculate the impacts of training on the nontruncated means of unemployment and employment duration as well as earnings as averages across our sample of treated persons.⁴⁰ A participation in short-term training reduces the expected unemployment duration by 53 days (−7.5%), increases the expected employment duration by 162 days (+19.1%) and reduces average daily earnings by €−0.46 (−1.0%). Long-term training increases unemployment duration by ninety-seven days (+15.7%), employment duration by 330 days (+34.9%) and daily earnings by €3.50 (+7.2%).

These calculations indicate that short-term training is likely cost effective, both from the taxpayer's and the participant's perspective. It reduces unemployment duration and the associated transfer payments by about €1,850. These savings in transfer payments exceed the course fees by a factor of three. In addition, short-term training positively affects the expected time in subsequent employment. In contrast, the picture is less clear for long-term training. With the additional benefit payments arising during the extra time in unemployment, a long-term training course causes about €9,000 of additional costs for the taxpayer compared with the situation of nonparticipation. Yet in view of the substantial positive effects on employment duration and the moderate earnings gains, it might be possible that in the long run also long-term training pays off from a fiscal point of view. As the benefits in terms of reduced unemployment insurance payments and higher tax revenues accumulate over time they may eventually exceed the initial investment of €9,000. Further, from the perspective of the participant, long-term training programs well seem to be attractive on average. Indeed, while unemployed participants receive relatively generous unemployment benefits. Once they have completed their long-term training course, on average, they quickly take up a job that is more stable and pays more than without training. A full assessment of the cost-effectiveness of both training programs according to a social cost-benefit criterion that considers the costs and benefits that accrue to taxpayers and participants alike, would require further assumptions and information, e.g., on discount rates and tax scales.

6. CONCLUSION

This study investigates and compares the dynamic causal effects of short-term, job search-oriented training and long-term, human capital intensive training schemes. Our empirical analysis uses rich administrative data for Germany, where both program types are used at the same time. We examine and compare the separate effects of both programs on unemployment and employment spells as well as daily earnings, taking into account the heterogeneity of training impacts according to the timing of participation during unemployment as well as across different subgroups of participants.

We find that participating in short-term training reduces the remaining time in unemployment and has moderate positive effects on subsequent job stability. Long-term training programs initially prolong the remaining time in unemployment, but once the scheduled program end

40. We compute the (nontruncated) expected treatment and nontreatment outcomes as follows: $E[T] = \int_0^{\bar{t}} S(r) dr + S(\bar{t})/\theta(\bar{t})$, with $S(t)$ the survivor function and $\theta(t)$ the hazard rate at $T = t$. This means, we extrapolate beyond truncation point \bar{t} assuming a constant hazard rate equal to the hazard rate at \bar{t} . We set \bar{t} to 1,460 days for unemployment duration, 1,825 days for employment duration, and €144 for daily earnings.

is reached participants exit to employment at a much faster rate than without training. Moreover, participants in long-term training benefit from substantially more stable employment spells and higher earnings.

Importantly, our findings point to the possibility of improving the efficiency of long-term training programs through a careful targeting. Specifically, we find that the opportunity cost of participating in terms of prolonged unemployment is lower for people with lower chances to exit unemployment on their own, e.g. long-term unemployed and low skilled. Persons without formal education degree and persons previously working in low- and medium-skilled manual occupations reap particularly high benefits with respect to earnings. Persons previously working in medium-skilled analytic and interactive occupations achieve substantial gains in terms of employment stability. Thus, long-term training programs seem well effective in supporting the occupational advancement of very heterogeneous groups of people, including those with generally weak labor market prospects. However, from a more narrow, fiscal perspective only the low-cost short-term training schemes are on average cost efficient in the short run.⁴¹ In light of the recent policy shifts favoring short, job search-oriented training programs a more balanced use of both types of training seems warranted. In fact, while the shorter programs may appeal for their cost efficiency the longer programs seem more effective in tackling structural deficits.

From a conceptual point of view, our study highlights that time is an important dimension of program evaluation not appropriately accounted for in conventional static and experimental research designs. A detailed analysis of the dynamics of program assignment and labor market outcomes, considering also impact heterogeneity allows a deeper understanding of how programs work. This knowledge is important for an optimal use of public employment services.

Acknowledgments. I am grateful to the editor Imran Rasul and three anonymous referees for their very helpful comments and suggestions. The article also benefited from comments by Christian Dustmann, Bernd Fitzenberger, Volker Grossmann, Krisztina Kis-Katos, Marie Paul, Jeff Smith, Gerard van den Berg, and participants at various conferences, workshops and seminars in Amsterdam, Freiburg, Graz, London, Mannheim, Milan, Nottingham, Nuremberg, Washington and Zurich. The article is part of the project “Employment effects of further training programs 2000-2002—An evaluation based on register data provided by the Institute of Employment Research (IAB)” (IAB project number 6-531.1A). Special thanks to Stefan Bender for his efforts to make the data available. Financial support by the IAB and the German Excellence Initiative is gratefully acknowledged.

SUPPLEMENTARY DATA

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

- ABBRING, J.H. (2008), “The Event-History Approach to Program Evaluation” , in Millimet, D.L., Smith, J.A. and Vytlačil, E. (eds), *Advances in Econometrics*, Volume 21: Modeling and Evaluating Treatment Effects in Econometrics (Amsterdam: Elsevier), 33–55.
- ABBRING, J.H. and HECKMAN, J.J. (2007), “Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation”, in Heckman, J. J. and Leamer, E. E. (eds), *Handbook of Econometrics*, Vol. 6B (Amsterdam: Elsevier), 5145–5303.
- ABBRING, J.H. and VAN DEN BERG, G.J. (2003a), “The Nonparametric Identification of Treatment Effects in Duration Models”, *Econometrica*, **71**, 1491–1517.
- ABBRING, J.H. and VAN DEN BERG, G.J. (2003b), “The Identifiability of the Mixed Proportional Hazards Competing Risks Model”, *Journal of the Royal Statistical Society, Series B*, **65**, 701–710.
- ABBRING, J.H. and VAN DEN BERG, G.J. (2004), “Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models”, *Empirical Economics*, **29**, 5–20.
- ABBRING, J.H. and VAN DEN BERG, G.J. (2005), “Social Experiments and Instrumental Variables with Duration Outcomes” (Tinbergen Institute Discussion Paper TI 2005-047/3).

41. All these conclusions hold with the caveat that general equilibrium effects are not accounted for.

- ACEMOGLU, D. and AUTOR, D. (2011), "Skills, Tasks and Technologies: Implications for Employment and Earnings", in Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*, Vol. 4B (Amsterdam: Elsevier), 1043–1166.
- ASHENFELTER, O. (1978), "Estimating the Effect of Training Programs on Earnings", *Review of Economics and Statistics*, **60**, 47–57.
- AUTOR, D.H., KATZ, L.F. and KEARNEY, M.S. (2008), "Trends in U.S. Wage Inequality: Revising the Revisionists", *Review of Economics and Statistics*, **90**, 300–323.
- BERGEMANN, A., FITZENBERGER, B. and SPECKESSER, S. (2009), "Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences", *Journal of Applied Econometrics*, **24**, 797–823.
- BLANK, R. (2002), "Evaluating Welfare Reform in the United States", *Journal of Economic Literature*, **40**, 1105–1166.
- BONNAL, L., FOUGÈRE, D. and SÉRANDON, A. (1997), "Evaluating the Impact of French Employment Policies on Individual Labour Market Histories", *Review of Economic Studies*, **64**, 683–713.
- BRINCH, C. (2007), "Nonparametric Identification of the Mixed Hazards Model with Time-Varying Covariates", *Econometric Theory*, **23**, 349–354.
- CARD, D., KLUVE, J. and WEBER, A. (2010), "Active Labour Market Policy Evaluations: A Meta-Analysis", *Economic Journal*, **120**, F452–F477.
- COUCH, K.A. (1992), "New Evidence on the Long-Term Effects of Employment Training Programs", *Journal of Labor Economics*, **10**, 380–388.
- CRÉPON, B., DEJEMEPPE, M. and GURGAND, M. (2005), "Counseling the Unemployed: Does It Lower Unemployment Duration and Recurrence?" (IZA Discussion Paper No. 1796).
- CRÉPON, B., FERRACCI, M. and FOUGÈRE, D. (2012), "Training the Unemployed in France: How Does It Affect Unemployment Duration and Recurrence?" forthcoming in *Annales d'Economie et de Statistique*.
- CRÉPON, B., FERRACCI, M., JOLIVET, G. and VAN DEN BERG, G. J. (2009), "Active Labor Market Policy Effects in a Dynamic Setting", *Journal of the European Economic Association*, **7**, 595–605.
- DE LUNA, X. and JOHANSSON, P. (2010), "Non-Parametric Inference for the Effect of a Treatment on Survival Times with Application in the Health and Social Sciences", *Journal of Statistical Planning and Inference*, **140**, 2122–2137.
- DONALD, S.G., GREEN, D.A. and PAARSCH, H.J. (2000), "Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates", *Review of Economic Studies*, **67**, 609–633.
- DUSTMANN, C., LUDSTECK, J. and SCHÖNBERG, U. (2009), "Revisiting the German Wage Structure", *Quarterly Journal of Economics*, **124**, 843–881.
- DYKE, A., HEINRICH, C.J., MUESER, P. R., TROSKE, K. R. and JEON, K.-S. (2006), "The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes", *Journal of Labor Economics*, **24**, 567–607.
- EBERWEIN, C., HAM, J.C. and LALONDE, R. J. (1997), "The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data", *Review of Economic Studies*, **64**, 655–682.
- ELSBY, M.W.L., HOBIJN, B. and ŞAHIN, A. (2010), "The Labor Market in the Great Recession", *Brookings Papers on Economic Activity*, Spring, 1–69.
- EUROPEAN COMMISSION (2002), "Taking Stock of Five Years of the European Employment Strategy", COM 416(01), Commission of the European Communities, Brussels.
- FITZENBERGER, B., OSIKOMINU, A. and VÖLTER, R. (2008), "Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany", *Annales d'Economie et de Statistique*, **91–92**, 321–355.
- FREDRIKSSON, P. and JOHANSSON, P. (2008), "Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data", *Journal of Business and Economic Statistics*, **26**, 435–445.
- GAURE, S., RØED, K., VAN DEN BERG, G. J. and ZHANG, T. (2010), "Estimation of Heterogeneous Treatment Effects on Hazard Rates" (IZA Discussion Paper No. 4794).
- GREENBERG, D.H., MICHALOPOULOS, C. and ROBINS, P.K. (2003), "A Meta-Analysis of Government Sponsored Training Programs", *Industrial and Labor Relations Review*, **57**, 31–53.
- GRITZ, M. (1993), "The Impact of Training on the Frequency and Duration of Employment", *Journal of Econometrics*, **57**, 21–51.
- HAM, J.C. and LALONDE, R.J. (1996), "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training", *Econometrica*, **64**, 175–205.
- HECKMAN, J.J., ICHIMURA, H. and TODD, P.E. (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *Review of Economic Studies*, **64**, 605–654.
- HECKMAN, J.J., LALONDE, R.J. and SMITH, J.A. (1999), "The Economics and Econometrics of Active Labor Market Programs", in Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*, Vol. 3A (Amsterdam: Elsevier), 1865–2097.
- HECKMAN, J.J. and SINGER, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, **52**, 271–320.
- HECKMAN, J.J. and SMITH, J.A. (1999), "The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies", *Economic Journal*, **109**, 313–348.
- HECKMAN, J.J., SMITH, J. and CLEMENTS, N. (1997), "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *Review of Economic Studies*, **64**, 487–535.
- HONORÉ, B.E. (1993), "Identification Results for Duration Models with Multiple Spells", *Review of Economic Studies*, **60**, 241–246.

- HORNY, G. and PICCHIO, M. (2010), "Identification of Lagged Duration Dependence in Multi-Spell Competing Risks Models", *Economics Letters*, **106**, 241–243.
- HOTZ, V.J., IMBENS, G.W. and KLERMAN, J.A. (2006), "Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Reanalysis of the California GAIN Program", *Journal of Labor Economics*, **24**, 521–566.
- HUJER, R., THOMSEN, S.L. and ZEISS, C. (2006a), "The Effects of Short-Term Training Measures on the Individual Unemployment Duration in West Germany" (ZEW Discussion Paper 06-65).
- HUJER, R., THOMSEN, S.L. and ZEISS, C. (2006b), "The Effects of Vocational Training Programmes on the Duration of Unemployment in Eastern Germany", *AStA Advances in Statistical Analysis*, **90**, 299–321.
- KAPLAN, E.L. and MEIER, P. (1958), "Nonparametric Estimation from Incomplete Observations", *Journal of the American Statistical Association*, **53**, 457–481.
- KLUVE, J. (2010), "The Effectiveness of European Active Labor Market Policy", *Labour Economics*, **17**, 904–918.
- LALIVE, R., VAN OURS, J.C. and ZWEIMÜLLER, J. (2008), "The Impact of Active Labour Market Programmes on the Duration of Unemployment in Switzerland", *Economic Journal*, **118**, 235–257.
- LECHNER, M., MIQUEL, R. and WUNSCH, C. (2011), "Long-Run Effects of Public Sector Sponsored Training in West Germany", *Journal of the European Economic Association*, **9**, 742–784.
- LOK, J. (2008), "Statistical Modeling of Causal Effects in Continuous Time", *Annals of Statistics*, **36**, 1464–1507.
- NEYMAN, J. (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9", edited and translated by Dabrowska, D. M. and Speed, T. P. (1990), *Statistical Science*, **5**, 465–472.
- OECD (2005, 2006), *OECD Employment Outlook*. Paris (various issues).
- RICHARDSON, K. and VAN DEN BERG, G.J. (2006), "Swedish Labor Market Training and the Duration of Unemployment" (IZA Discussion Paper No. 2314).
- ROBINS, J.M. (1998), "Structural Nested Failure Time Models", in Armitage, P. and Colton, T. (eds), *The Encyclopedia of Biostatistics* (Chichester: John Wiley and Sons), 4372–4389.
- ROY, A.D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, **3**, 135–146.
- RUBIN, D.B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, **66**, 688–701.
- SIANESI, B. (2004), "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s", *Review of Economics and Statistics*, **86**, 133–155.
- SIANESI, B. (2008), "Differential Effects of Active Labour Market Programs for the Unemployed", *Labour Economics*, **15**, 370–399.
- SPITZ-OENER, A. (2006), "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure", *Journal of Labor Economics*, **24**, 235–270.
- VAN DER VAART, A. (1996), "Efficient Maximum Likelihood Estimation in Semiparametric Mixture Models", *Annals of Statistics*, **24**, 862–878.
- WALLER, M. (2009), "Many Dropouts? Never Mind!—Employment Prospects of Dropouts from Training Programs" (CDSE Mannheim Discussion Paper No. 63).
- WEBER, A. and HOFER, H. (2004a), "Are Job Search Programs a Promising Tool? A Microeconomic Evaluation for Austria" (IZA Discussion Paper No. 1075).
- WEBER, A. and HOFER, H. (2004b), "Employment Effects of Early Interventions on Job Search Programs" (IZA Discussion Paper No. 1076).